NEURAL NETWORK-BASED METAMODELLING APPROACH FOR ESTIMATION OF AIR POLLUTANT PROFILES

HERMAN WAHID

DOCTOR OF PHILOSOPHY

UNIVERSITY OF TECHNOLOGY, SYDNEY

2013

NEURAL NETWORK-BASED METAMODELLING APPROACH FOR ESTIMATION OF AIR POLLUTANT PROFILES

By

HERMAN WAHID

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

Faculty of Engineering
University of Technology, Sydney

February 2013

CERTIFICATE OF AUTHORSHIP/ ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a similar degree nor has it been submitted as part of requirements for any other degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are referenced in the thesis.

February 2013

....

Herman Wahid

This thesis is especially dedicated to my dearest father, mother, wife and family for their love, blessing and encouragement ...

ACKNOWLEDGMENT

First of all, I am greatly indebted to God in His mercy and blessing for making this project successful. This research project was carried out in the Faculty of Engineering, University of Technology Sydney between August 2009 and September 2012. I gratefully acknowledge the financial support for this research by the Malaysia Government and Universiti Teknologi Malaysia (UTM) Scholarship.

I would like to express my deepest gratitude and appreciation to my honourable supervisor, Associate Professor Dr. Quang Ha for his continuous guidance, encouragement, committed support, timely reviews and invaluable advice throughout the duration of this research project.

I would like to thank two atmospheric scientists, Dr. Hiep Duc from Office of Environment and Heritage New South Wales, and Dr. Merched Azzi from CSIRO Energy Technology New South Wales, for their assistance with air quality modelling related theories and their constructive comments with respect to this research project. Also thanks to Dr. Hiep Duc and his division team members for guiding and providing me with the valuable knowledge of air quality measurement methodologies for the collection of the research dataset.

My gratitude is also extended to Dr. Guenter A Plum for editing my thesis draft. Dr. Plum is an Australian academic thesis editor and proofreader, and has more than ten years experience in academic thesis editing, academic thesis proofreading and academic dissertation editing and proofing with clients in Australia, Asia, and around the world. Dr. Plum has worked in universities in Australia and Hong Kong as well as in the computer industry. Dr. Plum's helpful suggestions and attention to detail have been integral to a timely completion of my thesis.

I would also like to extend my thanks and best wishes to all students in my research group, especially to Vahid Vakiloroaya and Raja M Taufika Raja Ismail for some technical supports during my study. Their presence and companionship made my study experience most comfortable and enjoyable.

Last but not least, I would like to express my sincere appreciation and gratitude to my parents, Wahid Moktar and Mazenah Sidik, for their patience, encouragement and support throughout the years of my study. I would like to dedicate this research to my beloved wife, Nazariah Jadon. Her constant empathy and consistent moral support have eased the tough times during my research.

ABSTRACT

The air quality system is a system characterised by non-linear, complex relationships. Among existing air pollutants, the ozone (O_3), known as a secondary pollutant gas, involves the most complex chemical reactions in its formation, whereby a number of factors can affect its concentration level. To assess the ozone concentration in a region, a measurement method can be implemented, albeit only at certain points in the region. Thus, a more complicated task is to define the spatial distribution of the ozone level across the region, in which the deterministic air quality model is often used by the authority. Nevertheless, simulation by using a deterministic model typically needs high computational requirements due to the nonlinear nature of chemical reactions involved in the model formulation, which is also subject to uncertainties. In the context of ozone as an air pollutant, the determination of the background ozone level (BOL), independent from human activities, is also important as it could represent one of reliable references to human health risk assessment. The concept of BOL may be easily understood, but practically, it is hard to distinguish between natural and anthropogenic effects. Apart from existing approaches to the BOL determination, a new quantisation method is presented in this work, by evaluating the relationship of ozone versus nitric oxide (O_3-NO) to estimate the BOL value, mainly by using night-time and early morning measurement data collected at the monitoring stations.

In this thesis, to deal with the challenging problem of air pollutant profile estimation, a metamodel approach is suggested to adequately approximate intrinsically non-linear and complex input-output relationships with significantly less computation. The intrinsic characteristics of the underlying physics are not assumed to be known, while the system's input and output behaviours remain essential. A considerable number of metamodels approach have been proposed in the literature, e.g. splines, neural networks, kriging and support vector machine. Here, the radial basis function neural network (RBFNN) is concerned as it is known to offer good estimation performance on accuracy, robustness, versatility, sample size, efficiency, and

simplicity as compared to other stochastic approaches. The development requirements are that the proposed metamodels should be capable of estimating the ozone profiles and its background level temporally and spatially with reasonably good accuracies, subject to satisfying some statistical criteria.

Academic contributions of this thesis include in a number of performance enhancements of the RBFNN algorithms. Generally, three difficulties involved in the network training, selection of radial basis centres, selection of the basis function variance (i.e. spread parameter), and training of network weights. The selection of those parameters is very crucial, as they directly affect the number of hidden neurons used and also the network overall performance. In this research, some improvements of the typical RBFNN algorithm (i.e. orthogonal least squares) are achieved. First, an adaptively-tuned spread parameter and a pruning algorithm to optimise the network's size are proposed. Next, a new approach for training the RBFNN is presented, which involves the forward selection method for selecting the radial basis centres. Also, a method for training the network output weights is developed, including some suggestions for estimation of the best possible values of the network parameters by considering the cross-validation approach. For applications, results show that the combination of the proposed paradigm could offer a sub-optimal solution of metamodelling development in the generic sense (by avoiding the iteration process) for a faster computation, which is essential in air pollutant profile estimation.

PUBLICATIONS

Journal Articles

1. Hiep Duc, Merched Azzi, Herman Wahid, Q.P. Ha, "Background ozone level in the Sydney basin: Assessment and trend analysis", *Journal of Climatology*, in Press, doi: 10.1002/joc.3595.
2. H. Wahid, Q.P. Ha, H. Duc, "New sampling scheme for neural network-based metamodelling with application to air pollutant estimation," *Gerontechnology*, vol. 11(2), 2012, pp. 336, doi:10.4017/gt.2012.11.02.325.00.
3. H. Wahid, Q.P. Ha, H. Duc, M. Azzi, "Meta-modelling approach for estimating the spatial distribution of air pollutant levels", *Journal of Applied Soft Computing* (revised and re-submitted).
4. Q.P. Ha, H. Wahid, H. Duc, M. Azzi, "Radial basis function neural network-based metamodelling for ozone spatial estimation," *IEEE Transactions on Neural Network and Learning Systems* (submitted).

Conference Proceedings

1. H. Wahid, Q.P. Ha, H. Duc, "New sampling scheme for neural network-based metamodelling with application to air pollutant estimation," *Proc. 8th World Conference of Gerontechnology & 29th International Symposium for Automation and Robotics in Construction (ISG-ISARC 2012)*, Eindhoven, Netherlands, June 26-29, 2012.
2. Herman Wahid, Q.P. Ha, Hiep Duc, Merched Azzi, "Estimation of background ozone temporal profiles using neural networks," *Proc. 3rd IEEE Int. Conf. on Intelligence Computing and Intelligent Systems (ICIS 2011)*, Guangzhou, China, 18-20 Nov 2011, pp. 292-297.
3. Herman Wahid, Q.P. Ha, Hiep Nguyen Duc, "Computational intelligence estimation of natural background ozone level and its distribution for air quality modelling and emission control," *Pro. 28th International Symposium on*

Automation and Robotics in Construction (ISARC 2011), Seoul, Korea, 29 Jun-2 Jul 2011, pp. 551-557.

4. Herman Wahid, Quang P. Ha, and Hiep Nguyen Duc, "A Metamodel for Background Ozone Level Using Radial Basis Function Neural Networks," *Pro. 11th International Symposium on Control, Automation, Robotics and Vision (ICARCV 2010)*, Singapore, 7-10 Dec 2010, pp. 958-963.

5. H. Wahid, Q. P. Ha, and H. Duc, "Adaptive Neural Network Metamodel for Short-term Prediction of Natural Ozone Level in an Urban Area," *Proc. Int. Conf. on Computing and Communication Technologies (2010 IEEE-RIVF)*, Hanoi, Vietnam, 1-4 Nov 2010, pp. 250-253.

Related publications

1. Herman Wahid, Hiep Nguyen Duc, and Quang P. Ha, "Radial Basis Function Neural Network Metamodelling for 2-D Resistivity Mapping," *Proc. 27th International Symposium on Automation and Robotics in Construction (ISARC 2010),* Bratislava, Slovakia, 25-27 Jun 2010, pp. 364-373.

2. H. Wahid, Q.P. Ha, and M.S. Mohamed Ali, "Optimally-Tuned Cascaded PID Control using Radial Basis Function Neural Network Metamodeling," *Proc. of the 3rd International Workshop on Artificial Intelligence in Science and Technology (AISAT'09)*, Hobart, Australia, 23-24 November 2009, paper S01.2.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figures **Page**

LIST OF ABBREVIATIONS

σ	-	RBF spread parameter
σ_c	-	Isotropic spread parameter for RBF centres selection
λ	-	Regularisation parameter
AQM	-	Air quality management
BOL	-	Background ozone level
CGS	-	Classical Gram-Schmidt
d_2	-	Index of agreement
DAQM	-	Deterministic air quality model
DOE	-	Design of experiments
EPA	-	Environment Protection Authority (in Australia)
FFD	-	Full factorial design
FS	-	Forward selection
GCV	-	Generalised cross-validation
GLS	-	Generalised least squares
GMR	-	Greater Metropolitan Region
LHD	-	Latin hypercube design
LOO-CV	-	Leave one out cross-validation
LS	-	Least squares
MAE	-	Mean absolute errors
MSE	-	Mean squared errors
NAAQS	-	National Ambient Air Quality Standards
NEPM	-	National Environment Protection Measures
NO	-	Nitric oxide
NO_2	-	Nitrogen dioxide
NO_x	-	Nitrogen oxides
O_3	-	Ozone
OLS	-	Orthogonal least squares
PM	-	Particulate matter
ppb	-	parts per billion

pphm	-	parts per hundred million
R^2	-	Determination coefficient
RBFNN	-	Radial basis function neural network
RMSE	-	Root mean square errors
SSE	-	Sum of squared errors
TEMP	-	Ambient temperature
US-EPA	-	U.S. Environmental Protection Agency
VOCs	-	Volatile organic compounds
WCD	-	Weighted clustering design
WDR	-	Wind direction
WLS	-	Weighted least squares
WSP	-	Wind speed